

Validation of ICD-10-CM Codes for Identifying Cases of Chlamydia and Gonorrhea

Yenling A. Ho, MPH¹, Saurabh Rahurkar, BDS, DrPH^{2,3}, Guoyu Tao, PhD⁴, Chirag G. Patel,
DC, MPH⁴, Janet N. Arno, MD^{5,6}, Jane Wang, PhD⁷, Andrea A. Broyles⁷, MPH,
Brian E. Dixon, MPA, PhD^{1,7}

¹Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, IN, USA; ²The
Ohio State University College of Medicine, Columbus, OH, USA; ³The Ohio State University
Center for the Advancement of Team Science, Analytics, and Systems Thinking, Columbus, OH,
USA; ⁴Centers for Disease Control and Prevention, Atlanta, GA, USA; ⁵Indiana University
School of Medicine, Indianapolis, IN, USA; ⁶Marion County Public Health Department,
Indianapolis, IN, USA; ⁷Regenstrief Institute, Indianapolis, IN, USA

Corresponding author and alternate corresponding author:

Yenling Andrew Ho, MPH

Department of Epidemiology

Indiana University (IU) Richard M. Fairbanks School of Public Health

Indianapolis, USA

Phone: 765-717-5198

Email: yeho@iu.edu

This is the author's manuscript of the article published in final edited form as:

Ho, Y. A., Rahurkar, S., Tao, G., Patel, C. G., Arno, J. N., Wang, J., Broyles, A. A., & Dixon, B. E. (2020). Validation of ICD-10-CM Codes for Identifying Cases of Chlamydia and Gonorrhea. Sexually Transmitted Diseases. <https://doi.org/10.1097/OLQ.0000000000001257>

Brian E. Dixon, MPA, PhD

Center for Biomedical Informatics

Regenstrief Institute

Indianapolis, IN 46202 USA

Conflicts of Interest and Source of Funding: The authors report no conflicts of interest. This work was supported by a contract from the U.S. Centers for Disease Control and Prevention [Contract number: 2017-Q-67119] and a grant from the U.S. National Library of Medicine [Grant number: T15LM012502]. The views expressed in this publication are those of the authors and do not necessarily reflect the position or policy of the Centers for Disease Control and Prevention, National Library of Medicine, or the United States government.

Acknowledgements: The authors would like to thank Ashley Wiensch, MPH for her help in coordinating the team's efforts.

Short Summary

A study about using ICD-10-CM codes to identify chlamydia and gonorrhea cases found that their sensitivity was low (<15%) but their specificity was high (99.9%)

ABSTRACT

Background: While researchers seek to use administrative health data to examine outcomes for individuals with sexually transmitted infections, the ICD-CM-10 codes used to identify persons with chlamydia and gonorrhea have not been validated. Objectives were to determine the validity of using ICD-10-CM codes to identify individuals with chlamydia and gonorrhea.

Methods: We utilized data from electronic health records gathered from public and private health systems from October 1, 2015 to December 31, 2016. Patients were included if they were aged 13-44 years and received either 1) laboratory testing for chlamydia or gonorrhea or 2) an ICD-10-CM diagnosis of chlamydia, gonorrhea, or an unspecified STI. To validate ICD-10-CM codes, we calculated positive and negative predictive values, sensitivity, and specificity based on the presence of a laboratory test result. We further examined the timing of clinical diagnosis relative to laboratory testing.

Results: The positive predictive values for chlamydia, gonorrhea, and unspecified STI ICD-10-CM codes were 87.6%, 85.0%, and 32.0%, respectively. Negative predictive values were high (>92%). Sensitivity for chlamydia diagnostic codes was 10.6% and gonorrhea was 9.7%. Specificity was 99.9% for both chlamydia and gonorrhea. The date of diagnosis occurred on or after the date of the laboratory result for 84.8% of persons with chlamydia, 91.9% for gonorrhea, and 23.5% for unspecified STI.

Conclusions: Disease specific ICD-10-CM codes accurately identify persons with chlamydia and gonorrhea. However, low sensitivities suggest that most individuals could not be identified in administrative data alone without laboratory test results.

Keywords: ICD-10-CM; Administrative Health Data; Chlamydia; Gonorrhea; Validation

INTRODUCTION

Administrative data are routinely collected healthcare claims and service delivery data that detail diagnoses and procedures as well as medication prescriptions for populations that interface with the health system. These data are a convenient source for public health researchers to assess incidence¹ and prevalence^{2,3}, quality of healthcare^{4,5}, health outcomes^{6,7}, patient safety⁸, and econometrics⁹, areas of particular interest as STI services move away from STI clinics to primary care providers¹⁰. Researchers increasingly seek to leverage large administrative databases, such as all-payer claims databases, for population-level studies. In these studies, specific health conditions or outcomes are typically identified using administrative data standards for clinical diagnoses, the International Classification of Diseases, Ninth and Tenth Revisions, Clinical Modification (ICD-9-CM and ICD-10-CM) codes as laboratory test results are not available. Examples include an analysis of chlamydia surveillance in British Columbia¹¹ as well as the Global Burden of Disease project¹², which uses ICD codes to estimate incidence and burden of many diseases.

Despite frequent use of ICD-9-CM and ICD-10-CM codes in studies, the validity of ICD codes for identification of conditions has not been well examined. While there exist studies which have tested the accuracy of ICD codes for conditions such as male infertility¹³, acute ischemic stroke¹⁴, and pneumonia¹⁵, evidence of the accuracy of ICD codes for sexually transmitted infections (STIs) remains largely unexamined. A recent systematic review¹⁶ found only two studies^{17,18} which assessed chlamydia and gonorrhea in the context of pelvic inflammatory disease (PID). These two studies reported a wide range of positive predictive values from 56% to 100% and only assessed ICD-9-CM codes as they were conducted before October 1, 2015, the mandatory transition date from ICD-9-CM to ICD-10-CM in the US.

Our objectives for this study were to assess 1) validity of using ICD-10-CM diagnostic codes to identify persons with chlamydia and gonorrhea; and 2) the timing of provider's ICD-10-CM diagnosis in relation to availability of laboratory testing.

MATERIALS AND METHODS

Study Design and Setting

We employed a retrospective cohort study of patients extracted from an existing STI testing registry from October 1, 2015 to December 31, 2016 described previously¹⁹. The registry consists of electronic health records (EHRs), including administrative and laboratory testing data, from the Indiana Network for Patient Care (INPC) linked to clinical (morbidity) and testing records from Indiana's largest public health program of individuals tested for chlamydia, gonorrhea, or syphilis from January 1, 2004 to December 31, 2016. Individuals were matched based on name, date of birth, social security number, race, and ethnicity. ICD-10-CM codes were only in INPC data whereas tests were contributed by both INPC and STI program data.

The INPC²⁰⁻²² is the nation's oldest and most comprehensive community health information exchange (HIE). The HIE network connects over 100 health care facilities including hospitals, physicians' practices, pharmacy networks, long-term post-acute care facilities, laboratories, and radiology centers across Indiana. The INPC is primarily used to support clinical care by making data available to clinicians, although it is also routinely used to support health services and public health research²³.

To evaluate whether ICD-10-CM codes for chlamydia and gonorrhea were valid, we compared ICD diagnoses from the INPC with laboratory testing data available from the STI testing registry. This study was approved by Indiana University's Institutional Review Board.

Cohort Assembly and Data Extraction

Because gonorrhea and chlamydia occurs most often among people at reproductive age²⁴, our study focused on patients aged 13 to 44 years. Patients were included if they were between 13 and 44 years old and either received an ICD-10-CM clinical diagnosis documented in their EHR or possessed a laboratory test result for chlamydia or gonorrhea. Individuals with unique episodes of care involving an ICD-10-CM code or positive laboratory result for an STI were classified as a *case* per usual public health surveillance practice. The ICD-10-CM codes utilized to identify STI cases were 1) chlamydia: A55, A56, A71, and A74.9; 2) gonorrhea: A54; and 3) unspecified STI: A64.

Per CDC de-duplication case definition²⁵, patients with multiple of the above ICD diagnoses for a given disease fewer than 30 days apart were defined as only one disease case. If a patient had multiple ICD diagnoses for a given disease >30 days apart, only the first disease case was selected and assessed for the study. Among patients with no documented ICD-10-CM diagnosis for chlamydia or gonorrhea during the study period but who possessed at least one positive laboratory test for one of these STIs, their first positive test result was selected and assessed for the study. Finally, among patients with no documented ICD-10-CM diagnosis for chlamydia or gonorrhea during the study period and who possessed no positive laboratory test result during the study period, their first negative test result was selected and assessed for the study.

Cohorts for each group of ICD-10-CM codes were extracted from the registry based on inclusion criteria and de-duplication rules. For each case, we extracted the subject identifier, ICD-10-CM diagnosis code, date of ICD diagnosis, laboratory test specimen collection date, and date of laboratory test result.

Assessment of Testing for Cases Diagnosed with an STI

We first assessed whether those cases involving an ICD-10-CM diagnosis for an STI also contained a laboratory test result for the same STI. Each case was linked with available laboratory test result(s) that were within 30 days of the date of ICD diagnosis.

Validation of ICD-based STIs

Case status was determined based on the combination of ICD-10-CM diagnosis plus laboratory test results. With respect to chlamydia and gonorrhea, true positive (TP) cases were defined as those with both a disease-specific ICD diagnosis and a corresponding positive laboratory test. False positive (FP) cases were defined as those with a disease-specific ICD diagnosis and a corresponding negative laboratory test. Individuals with a disease-specific ICD diagnosis without a corresponding laboratory test were removed from the ICD validation assessment as their disease status could not be confirmed.

False negative (FN) cases were defined as individuals with a positive laboratory test result but no corresponding, disease-specific ICD diagnosis. True negative (TN) cases were defined as individuals with a negative laboratory test result and who did not have an ICD diagnosis. The algorithm to classify case status is depicted in **Figure 1**.

For the assessment of cases involving the unspecified STI diagnostic code (A64), TP cases were defined as those with any STI positive laboratory test, STI tests are shown in **Table 3**. FP cases were defined as those with either all STI negative results or no laboratory test. We included unspecified STI cases with no laboratory test to determine if STI-related visual examinations may be documented in the medical chart. However, the chart review found no visual examinations were documented in the INPC. FN cases were not defined for this cohort because it could not be

determined *a priori* whether a patient with a positive STI laboratory test would have either received an unspecified STI diagnosis or the corresponding STI-specific ICD code. TN cases were not classified as it cannot be determined whether a provider was testing for specific STIs or testing because the STI condition was unknown.

Due to the potential number of STI tests within 30 days of an A64 diagnosis, the corresponding laboratory test(s) at a specific date was determined based on several criteria. First, if the A64 diagnosis date and laboratory test specimen collection or result date were matched, then the test on the specimen collection or result date was considered the corresponding laboratory test. Second, if a case had multiple specimen collection dates, the laboratory test with the collection date closest to the ICD diagnosis date was determined to be the corresponding laboratory test. Finally, if two laboratory tests had collection dates the same number of days from the date of the ICD diagnosis, but one specimen collection date is before and the other is after the date of diagnosis, the laboratory test with the collection date before the date of the ICD diagnosis was selected. On the date of the corresponding laboratory test, multiple laboratory tests could be ordered.

Medical Chart Reviews for Data Quality

For data quality purposes, we conducted chart reviews to check whether ICD diagnoses or laboratory test results missing from designated fields in the EHR might be present in other parts of the patient record, such as physician notes. Chart reviews included all cases where the case possessed an ICD diagnosis of chlamydia or gonorrhea with no corresponding positive laboratory test result.

Additional chart reviews were conducted among 300 randomly selected cases with an ICD diagnosis of A64 as well as 300 cases with a positive laboratory test result for either STI but no disease-specific ICD diagnosis. No additional ICD diagnoses or laboratory test results were found during the chart reviews; therefore, all STI laboratory testing procedures and results were captured within each patient's EHR.

Assessment of Clinical Diagnosis Relevant to Laboratory Testing

To assess provider ICD diagnostic behavior, we compared the date of the laboratory test result to the date of ICD diagnosis for those cases which possessed both an ICD diagnostic code and a laboratory test result. If a patient had multiple test result dates, the date closest to the ICD diagnosis date was selected. Further, if two dates had the same number of days from the date of the ICD diagnosis, but one test result date is before and the other is after the date of diagnosis, the test result date before the date of the ICD diagnosis was selected.

Statistical Analyses

Descriptive statistics were calculated to assess whether those cases with an ICD-10-CM diagnostic code possessed a corresponding laboratory test result within 30 days. The proportion of cases with a corresponding positive test, corresponding negative test, or missing laboratory testing information were calculated using only those cases with an ICD-10-CM diagnostic code. For patients with either positive or negative laboratory test results, using the case status definitions, we further calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for chlamydia and gonorrhea (disease-specific ICD cases). Sensitivity denotes the proportion of cases which are TP compared with all cases with disease. Specificity denotes the proportion of cases which are TN compared to all cases without disease. PPV denotes the

proportion of cases which are TP cases compared to cases with a disease-specific ICD diagnosis. NPV denotes the proportion of cases which are TN compared cases without a disease-specific ICD diagnosis. Confidence intervals were calculated by exact binomial method for each sensitivity, specificity, PPV, and NPV. For the unspecified STI cohort, because of the in-depth chart review process, we decided to only include those cases which were chart reviewed in the analysis. To provide unbiased estimates that are representative of the population of the INPC, all rates, estimates, and confidence intervals (95% CI) for the unspecified STI cohort were weighted based on the sampling methodology.

To examine diagnosis dates relative to laboratory testing dates, we calculated the difference between the date of laboratory test result and the date of ICD diagnosis for TP and FP cases. Negative numbers indicate the date of diagnosis occurred before the laboratory result date, “0” indicates that the date of ICD diagnosis was the same as the laboratory result date, and positive numbers indicate the date of diagnosis occurred after the laboratory result date. All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Study Population

Over an 18-month period, a total of 1,489,598 unique patients, between 13 and 44 years old, had at least one encounter in the INPC. Of these, a total of 238,876 individuals (16.0%) were either tested for chlamydia or gonorrhea, or diagnosed with an ICD-10-CM code of interest; 81,572 patients received an ICD diagnosis or were tested for chlamydia, 79,640 for gonorrhea, and 603 for an unspecified STI.

Assessment of Laboratory Testing for Cases Diagnosed with an STI

Among these cohorts, 914 patients received an ICD diagnosis of chlamydia, 349 for gonorrhea, and 603 for an unspecified STI. Information on laboratory testing among those cases with an ICD-10-CM diagnosis is summarized in **Table 1**. The proportion of diagnosed chlamydia or gonorrhea cases with a corresponding positive laboratory test result was 81.9% and 78.2% respectively. Of unspecified STI cases, 32.0% possessed a corresponding positive laboratory test for any STI and 24.7% of cases possessed a corresponding positive laboratory test for chlamydia or gonorrhea.

Validation of ICD-based STIs

We identified 749 TP, 106 FP, 6306 FN, and 74,352 TN for chlamydia, 273 TP, 48 FP, 2552 FN, and 76,739 TN cases for gonorrhea, and 193 TP and 410 FP cases for unspecified STI. All case statuses and validity measures are depicted in **Table 2**. Positive predictive values varied among each disease cohort with 87.6% for chlamydia, 85.0% for gonorrhea, and 32.0% for unspecified STI. Negative predictive values were high with 92.2% for chlamydia and 96.8% for gonorrhea. Sensitivities were low with 10.6% for chlamydia and 9.7% for gonorrhea. Specificities were 99.9% for both chlamydia and gonorrhea.

A total of 603 patients received an ICD diagnosis of an unspecified STI. Just 1.3% (8/603) of these unspecified STI cases received a disease-specific ICD diagnosis in addition to the non-specific STI diagnosis (4 chlamydia, 2 chlamydia and gonorrhea, and 2 bacterial vaginosis). The laboratory test combinations ordered by providers are summarized in **Table 3**. Only 32.0% (193/603) were found to have at least one corresponding positive laboratory test result. Among the 603 cases, 82.6% (498/603) and 81.3% (490/603) received a test for chlamydia and/or gonorrhea,

respectively. Of the cases who received testing, only 21.5% (107/498) tested positive for chlamydia and 15.5% (76/490) for gonorrhea.

Assessment of Clinical Diagnosis in Relation to Laboratory Testing

Information on the date of STI diagnosis in relation to the date of laboratory test results is presented in **Figure 2**. The proportion of cases for which the date of diagnosis occurred on or after the date of laboratory result was 84.8% for chlamydia, 91.9% for gonorrhea, and 23.5% for unspecified STI. When the date of ICD diagnosis for chlamydia or gonorrhea occurred at or after the date of laboratory result, the ICD diagnosis was more likely to be TP (98.4% and 91.9%, respectively). When the date of ICD diagnosis for chlamydia or gonorrhea occurred prior to the date of laboratory result, the ICD diagnosis was more likely to be FP (89.5% and 92.3% FP, respectively).

DISCUSSION

Using a population-level cohort drawn from linked electronic health records with STI morbidity records, we found that most disease-specific ICD-10-CM codes for chlamydia and gonorrhea accurately identified clinically diagnosed patients. The PPVs observed in this study using retrospective chart review supports the use of disease-specific ICD-10-CM codes as a valid method to identify cases of chlamydia and gonorrhea for secondary analyses, including health services research. However, low sensitivities and patterns of the date of diagnosis occurring before the date of laboratory testing observed in this study warrant caution for those who desire to use ICD-10-CM codes to conduct chlamydia and gonorrhea-related health services research.

Using a STI registry composed of electronic health records pooled from multiple health systems and a large public health STI surveillance program, we determined that PPVs for

chlamydia and gonorrhea were 87.6% and 85.0%, respectively. The PPVs were likely affected by presumptive diagnoses by clinicians based on the comparison of ICD diagnosis dates to laboratory test result dates. Results showed that, especially among FP cases, some clinicians were diagnosing patients without a laboratory test or before the test results were available. However, the high rates of TP cases make ICD-10-CM codes a reasonable method to identify cases for those in public health programs who wish to examine STI treatment and outcomes.

Sensitivities for both STIs examined were low: chlamydia 10.6% and gonorrhea 9.7%. This suggests low usage of ICD-10-CM codes by clinicians when STI cases were identified. Compared to the 1,215 STI cases that possessed an ICD-10-CM code, we identified 8,858 patients with a positive confirmatory laboratory test for an STI but no clinical diagnosis code in the medical record. While some cases were diagnosed in a public health clinic that does not bill third-parties for reimbursement, most cases were identified from records in private health care clinics, hospitals, or emergency departments. The high number of cases without an ICD diagnosis is likely due to screening efforts by clinicians to test young women (aged 16-24 years) for chlamydia, a Healthcare Effectiveness Data and Information Set (HEDIS) performance metric in the U.S.²⁶ Additionally, another possible explanation could be that patients are receiving encounter-related ICD-10-CM codes (e.g. “general adult medical examination” or “encounter for gynecological examination”) rather than clinical diagnosis codes. Therefore, using a diagnostic code only approach to estimate chlamydia and gonorrhea prevalence or incidence is likely to underestimate true prevalence or incidence of disease and thus should not be utilized for public health surveillance. Instead, public health authorities should use administrative data in combination with laboratory-based reporting, electronic systems where possible, for surveillance efforts which are generally required under state laws in the United States^{27,28}.

We further examined individuals who were diagnosed with the non-specific ICD-10-CM code for an STI (A64). The A64 code was found to have a low PPV of 32.0% with clinicians testing for 7 different STIs with 24 unique testing combinations. Although we identified some positive chlamydia and gonorrhea cases in this cohort, incorporating the A64 code with the specific ICD codes to identify cases will result in an overall lower PPV due to the number of FPs and other STIs potentially indicated by this code.

This analysis has several limitations. The first is that specific codes may have been under-utilized because specific diagnosis is usually made after the encounter, and treatment may not be associated with an ICD-10 code. Second, some cases may have received an ICD diagnosis or a laboratory test result, but these data may have fallen outside of the end date of our data. Therefore, we may have misclassified some cases. However, based on our assessment of the time of ICD-10-CM diagnosis compared to laboratory test result dates, the misclassification should be minimal. Third, we were unable to assess the added value of encounter-related ICD-10 codes, the effect of reason for visit, CPT codes or clinical location of the encounter because they were not available in our registry. Additionally, the study dates may have contributed to the low usage of disease-specific ICD codes as clinicians may not have acclimated to utilizing ICD-10-CM codes which were implemented in the United States on October 15, 2015. Finally, our findings might not be generalizable outside of Indiana given that all providers were in a single state, although we did utilize data from multiple health systems of variable size and complexity.

To our knowledge, this is the first study which attempted to validate ICD-10-CM codes for identifying cases of chlamydia and gonorrhea. This study was conducted using a population-level cohort with EHR chart reviews from a large community-wide HIE network linked with records from a regional public health STI program. Another strength was that we were able to complete

chart reviews of all patients in the INPC who had an ICD diagnosis of chlamydia or gonorrhea without positive laboratory test result, although true disease status could not be confirmed in all cases rather only in subset of sampled cases.

In conclusion, there should be caution in using ICD-10-CM codes to identify cases of chlamydia or gonorrhea. Despite high PPVs for chlamydia and gonorrhea ICD-10-CM codes, these codes only identify a meager portion of patients with chlamydia or gonorrhea infection. Accuracy of ICD-10-CM codes were further associated with whether the date of diagnosis was on or after the date of laboratory testing. In addition, an unspecified STI ICD-10 code might identify some chlamydia or gonorrhea cases, but its PPVs for chlamydia or gonorrhea infection were low. Our results suggest that gonorrhea and chlamydia cases are better identified by laboratory test results alone or in combination with ICD-10-CM codes rather than only ICD-10 codes in administrative claims data.

REFERENCES

1. Insinga RP, Itzler RF, Pellissier JM, Saddier P, Nikas AA. The incidence of herpes zoster in a United States administrative database. *J Gen Intern Med.* 2005;20(8):748-753.
2. Faught E, Richman J, Martin R, et al. Incidence and prevalence of epilepsy among older US Medicare beneficiaries. *Neurology.* 2012;78(7):448-453.
3. Dellon ES, Jensen ET, Martin CF, Shaheen NJ, Kappelman MD. Prevalence of Eosinophilic Esophagitis in the United States. *Clin Gastroenterol Hepatol.* 2014;12(4):589-596.e581.
4. Iezzoni LI. Assessing Quality Using Administrative Data. *Ann Intern Med.* 1997;127(8_Part_2):666-674.
5. Kerr EA, McGlynn EA, Van Vorst KA, Wickstrom SL. Measuring Antidepressant Prescribing Practice in a Health Care System Using Administrative Data: Implications for Quality Measurement and Improvement. *The Joint Commission Journal on Quality Improvement.* 2000;26(4):203-216.
6. Cram P, Lu X, Kates SL, Singh JA, Li Y, Wolf BR. Total Knee Arthroplasty Volume, Utilization, and Outcomes Among Medicare Beneficiaries, 1991-2010 Knee Arthroplasty Volume, Use, and Outcomes. *JAMA.* 2012;308(12):1227-1236.
7. Mitchell JB, Bubolz T, Paul JE, et al. Using Medicare claims for outcomes research. *Med Care.* 1994;32(7 Suppl):Js38-51.
8. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care.* 2003;12 Suppl 2(Suppl 2):ii58-ii63.

9. Robinson RL, Birnbaum HG, Morley MA, Sisitsky T, Greenberg PE, Claxton AJ. Economic cost and epidemiological characteristics of patients with fibromyalgia claims. *J Rheumatol*. 2003;30(6):1318-1325.
10. Batteiger TA, Dixon BE, Wang J, et al. Where Do People Go for Gonorrhea and Chlamydia Tests: A Cross-Sectional View of the Central Indiana Population, 2003–2014. *Sex Transm Dis*. 2019;46(2):132-136.
11. Rekart ML, Gilbert M, Meza R, et al. Chlamydia public health programs and the epidemiology of pelvic inflammatory disease and ectopic pregnancy. *J Infect Dis*. 2013;207(1):30-38.
12. Mathers CD, Stein C, Ma Fat D, et al. Global Burden of Disease 2000: Version 2 methods and results. *World Health Organization, Geneva*. 2002.
13. Khandwala YS, Zhang CA, Li S, Cullen MR, Eisenberg ML. Validity of Claims Data for the Identification of Male Infertility. *Curr Urol Rep*. 2017;18(9):68.
14. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. *Stroke*. 1998;29(8):1602-1604.
15. Barber C, Lacaille D, Fortin PR. Systematic review of validation studies of the use of administrative data to identify serious infections. *Arthritis Care Res*. 2013;65(8):1343-1357.
16. Dixon BE, Rahrkar S, Ho Y, Arno JN. Reliability of administrative data to identify sexually transmitted infections for population health: a systematic review. *BMJ Health Care Inform*. 2019;26(1):e100074.

17. Satterwhite CL, Yu O, Raebel MA, et al. Detection of Pelvic Inflammatory Disease: Development of an Automated Case-Finding Algorithm Using Administrative Data. *Infect Dis Obstet Gynecol*. 2011;2011:7.
18. Ratelle S, Yokoe D, Blejan C, et al. Predictive value of clinical diagnostic codes for the CDC case definition of pelvic inflammatory disease (PID): implications for surveillance. *Sex Transm Dis*. 2003;30(11):866-870.
19. Dixon BE, Tao G, Wang J, et al. An Integrated Surveillance System to Examine Testing, Services, and Outcomes for Sexually Transmitted Diseases. *Stud Health Technol Inform*. 2017;245:361-365.
20. Biondich PG, Grannis SJ. The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. *J Public Health Manag Pract*. 2004;Suppl:S81-86.
21. McDonald CJ, Overhage JM, Barnes M, et al. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Aff (Millwood)*. 2005;24(5):1214-1220.
22. Overhage JM, Tierney WM, McDonald CJ. Design and implementation of the Indianapolis Network for Patient Care and Research. *Bull Med Libr Assoc*. 1995;83(1):48-56.
23. Dixon BE, Whipple EC, Lajiness JM, Murray MD. Utilizing an integrated infrastructure for outcomes research: a systematic review. *Health Info Libr J*. 2016;33(1):7-32.
24. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2018*. Atlanta: U.S. Department of Health and Human Services; 2019.

25. Centers for Disease Control and Prevention. De-Duplication Guidance for Gonorrhea and Chlamydia Laboratory Reports. Centers for Disease Control and Prevention. <https://www.cdc.gov/std/laboratory/de-duplication-guidance-june2016.pdf>. Published 2016. Accessed May 5th, 2019.
26. National Committee for Quality Assurance. Chlamydia Screening in Women (CHL). NCQA. HEDIS Measures and Technical Resources Web site. <https://www.ncqa.org/hedis/measures/chlamydia-screening-in-women/>. Published 2019. Accessed Oct 9, 2019.
27. Black J, Hulkower R, Suarez W, Patel S, Elliott B. Public Health Surveillance: Electronic Reporting as a Point of Reference. *J Law Med Ethics*. 2019;47(2_suppl):19-22.
28. Doyle TJ, Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. *Am J Epidemiol*. 2002;155(9):866-874.

Figure Legends

Figure 1. Algorithm used to classify chlamydia and gonorrhea cases based on ICD-10-CM diagnosis and laboratory testing results.

Figure 2. ICD-10-CM diagnosis dates compared to laboratory test result dates in days. Negative values indicate that the ICD diagnosis came before laboratory test result date.

ACCEPTED

Figure 1

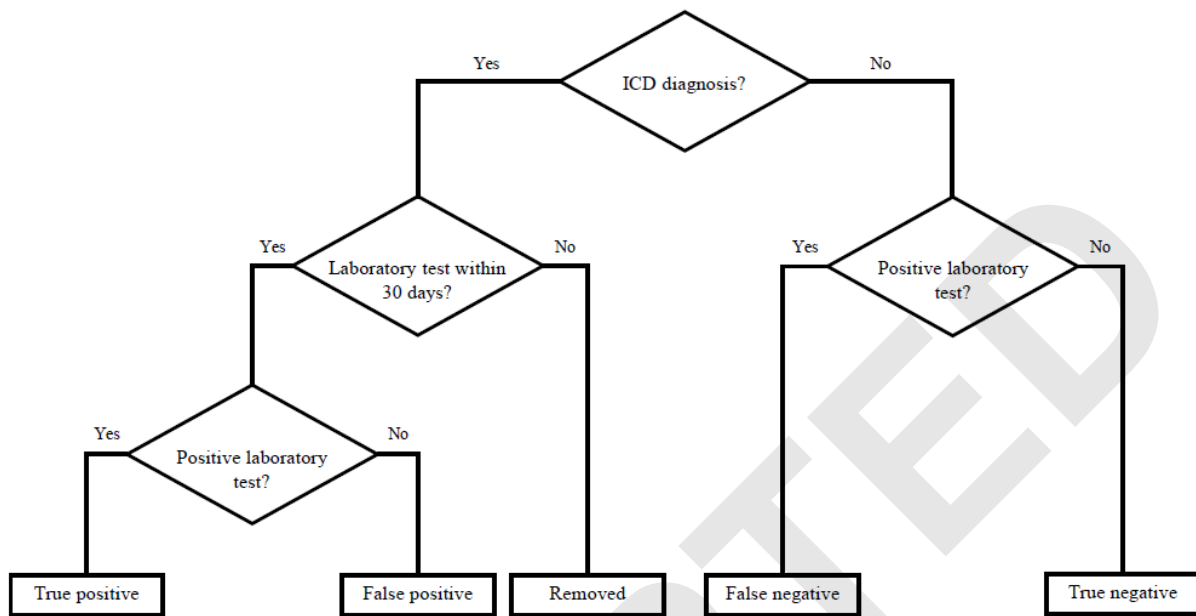


Figure 2

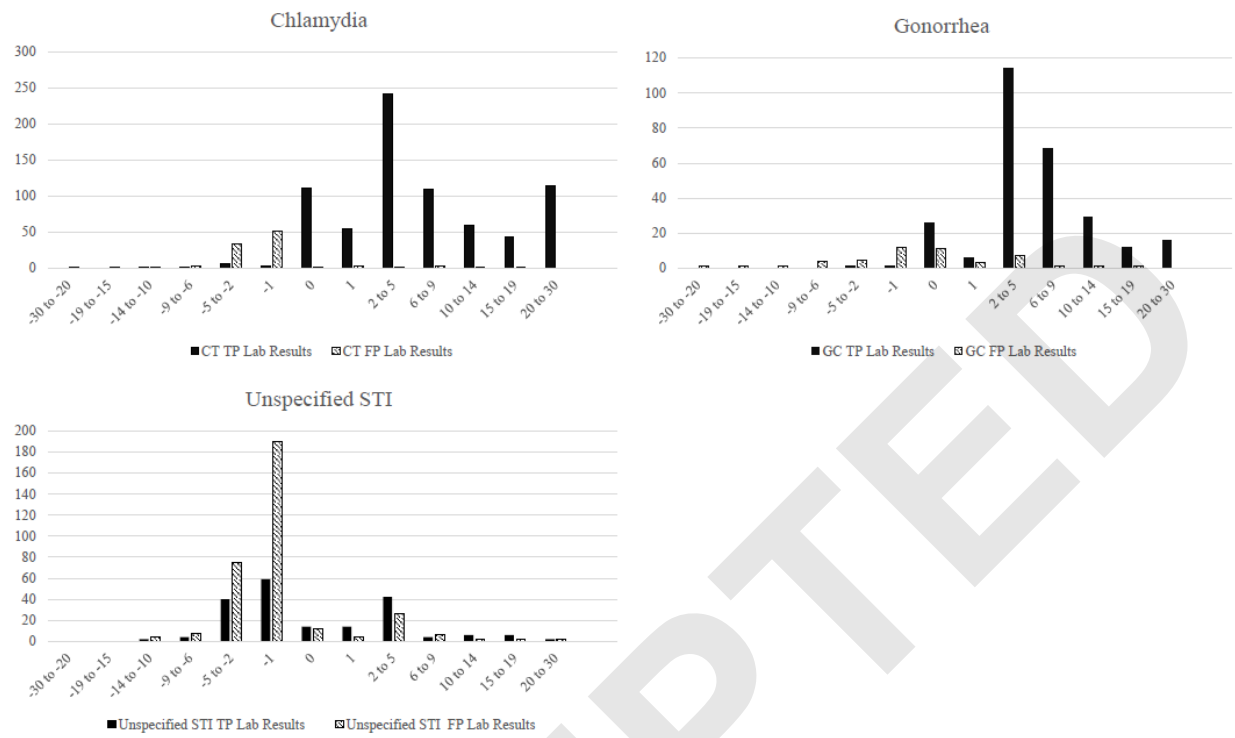


Table 1. Chlamydia, gonorrhea, and unspecified STI laboratory test results by associated ICD diagnosis code.

ICD Diagnosis	Laboratory Testing and Results			
	Positive (% among ICD)	Negative (% among ICD)	No Testing (% among ICD)	Total
Chlamydia	749 (81.9)	106 (11.6)	59 (6.5)	914
Gonorrhea	273 (78.2)	48 (13.8)	28 (8.0)	349
Unspecified STI (A64)	193 (32.0)	331 (54.9)	79 (13.1)	603

Unspecified STI (A64) values are weighted for the sampling methodology. Laboratory⁺ indicates presence of a positive laboratory test. Laboratory tests used to determine laboratory testing for unspecified STI are depicted in Table 3.

Table 2. Sensitivity, specificity, PPV, and NPV by phenotype.

Disease	TP	FP	FN	TN	Total	PPV	NPV	Sensitivity	Specificity
	(n)	(n)	(n)	(n)	(n)	%	%	y %	y %
						(95% CI)	(95% CI)	(95% CI)	(95% CI)
Chlamydia	74	10	630	7435	81513	87.6	92.2	10.6	99.9
	9	6	6	2		(87.6-87.6)	(92.2-92.2)	(10.6-10.6)	(99.9-99.9)
Gonorrhea	27	48	255	7673	79612	85.0	96.8	9.7	99.9
	3		2	9		(85.0-85.1)	(96.8-96.8)	(9.6-9.7)	(99.9-99.9)
Unspecified STI (A64)	19	41	---	---	603	32.0	---	---	---
	3	0				(32.0-32.1)			

Unspecified STI (A64) values are weighted for the sampling methodology. All 95% CI bounds were within 0.01% of the value. Laboratory tests used to determine TP and FP for unspecified STI are depicted in Table 3.

Table 3. Combination of Lab Orders Associated with the Diagnosis of Unspecified STI (A64).

Testing Combinations	Counts
CT	8
GC	2
SYPH	4
TRICH	8
HERP	2
CT/GC	175
CT/GC/SYPH	31
CT/GC/HIV	12
CT/GC/TRICH	131
CT/GC/HERPES	4
CT/GC/SYPH/HERP	4
CT/GC/SYPH/HIV	60
CT/GC/TRICH/HERP	2
CT/GC/HIV/TRICH	10
CT/GC/SYPH/TRICH	12
CT/SYPH/HIV/TRICH	2
CT/GC/SYPH/HIV/HERP	4
CT/GC/TRICH/GARD/CAND	18
CT/GC/SYPH/HIV/TRICH	20
CT/GC/SYPH/HIV/HERP	2

CT/GC/HIV/TRICH/GARD/CAND	4
SYPH/HIV	4
SYPH/HIV/HERP	2
TRICH/GARD/CAND	4
None	79

CT = Chlamydia; GC = Gonorrhea; SYPH = Syphilis; TRICH = Trichomoniasis; HERP = Herpes; Gard = Gardnerella vaginalis; CAND = Candidiasis